OQRE

Office de
la qualité et
de la responsabilité
en éducation

Le Test provincial de compétences linguistiques administré en octobre 2002 : Document technique traitant des résultats des processus de fiabilité et de validité

Le Test provincial de compétences linguistiques administré en octobre 2002 : Document technique traitant des résultats des processus de fiabilité et de validité

Le présent document technique vient compléter le document intitulé « *Le Test provincial de compétences linguistiques administré en octobre 2002 : Sommaire des mesures de contrôle de la qualité et des processus de fiabilité et de validité* » publié par l'Office de la qualité et de la responsabilité en éducation (OQRE). En plus des renseignements fournis dans le sommaire, le présent document offre des résultats détaillés sur la fiabilité et la validité du Test provincial de compétences linguistiques (TPCL) administré en octobre 2002. Nous présumons que, contrairement aux personnes à qui s'adressait le sommaire, les personnes qui liront le présent document ont une certaine connaissance des thèmes avancés en mesure et en statistiques.

A. Mesures de contrôle de la qualité visant à assurer l'uniformité de la notation

Avant la notation du TPCL, tous les travaux d'élèves reçus des écoles ont été mélangés, pour être distribués aux correctrices et correcteurs de façon aléatoire. Cela nous a permis de nous assurer qu'aucune personne ne recevait un nombre disproportionné de cahiers provenant d'une même école. De plus, *tous* les travaux d'élèves utilisés lors de la formation des correctrices et correcteurs et au cours de la notation même ont été *notés de façon anonyme*. La notation anonyme veut dire qu'aucun renseignement contenu dans les travaux des élèves ne permet à une correctrice ou un correcteur d'identifier une ou un élève, une école ou un conseil scolaire.

Nous pouvons assurer que les cahiers des élèves seront notés de manière exacte en procurant une formation appropriée aux correctrices et correcteurs. Tout au long de la période de notation du TPCL administré en octobre 2002, l'OQRE a suivi différentes étapes afin de s'assurer que les correctrices et correcteurs notaient les travaux d'élèves de manière semblable. Entre autres, l'OQRE a utilisé du matériel de formation, un test préparatoire de notation, la notation en chaîne, la notation en équipe et la notation double. Chacune de ces étapes est décrite dans le sommaire, affiché dans le site Web à l'adresse www.oqre.on.ca.

B. Fiabilité

Durant la notation du TPCL administré en octobre 2002, des cahiers d'élèves choisis de façon aléatoire ont été retournés à la pile de cahiers à noter afin qu'ils soient notés une deuxième fois. Ces cahiers, environ 900 appartenant à des élèves de langue française et 900 à des élèves de langue anglaise, représentaient une gamme étendue du rendement des élèves. Suivant la notation, les données tirées de la première et de la seconde notation ont été comparées afin d'obtenir un indice de la fiabilité interévaluateurs (entente avec les correctrices et correcteurs).

Une corrélation a été établie entre les notes totales de chaque élève provenant de la première et de la seconde notation, afin d'obtenir un indice de la fiabilité interévaluateurs. Les résultats obtenus par rapport à la fiabilité interévaluateurs indiquent que les cœfficients globaux de fiabilité dépassent 0,90 pour les deux composantes du test et pour les deux langues d'enseignement. Plus précisément, pour la composante lecture, le cœfficient de fiabilité interévaluateurs était de 0,93 (n = 875) pour les tests de langue française et de 0,99 (n = 882) pour les tests de langue anglaise. Pour la composante écriture, le cœfficient de fiabilité interévaluateurs était de 0,93 (n = 898) pour les cahiers de langue française et de 0,94 (n = 900) pour les cahiers de langue anglaise. Ces résultats indiquent que les réponses des élèves sont notées de manière uniforme, que les notes sont fiables et que les correctrices et correcteurs appliquent les grilles de notation de manière uniforme.

En ce qui a trait aux évaluations à grande échelle, les cœfficients de fiabilité interévaluateurs se situent habituellement entre 0,85 et 0,99 (consulter les documents '*Arizona's Instrument to Measure Standards' (AIMS)' Guide to Understanding AIMS Results; Massachusetts Comprehensive Assessment System: 1998 Technical Report* (1999); *Washington Assessment of Student Learning: Grade 10, 1999 Technical Report*, 2001). Ensemble, les résultats du TPCL administré en octobre 2002, sont conformes aux normes établies en matière de fiabilité interévaluateurs par rapport à la notation d'une évaluation à grande échelle.

C. Validité

Représentation du contenu du test

Pour assurer la validité du TPCL, il est important d'assurer que les items et les tâches sont tirés du programme-cadre du curriculum sur lequel le test est basé. Tous les types de textes de lecture et toutes les tâches d'écriture du TPCL sont basés sur les connaissances et les habiletés traitées dans *Le curriculum de l'Ontario* pour toutes les matières jusqu'à la fin de la 9^{e} année. Les élèves doivent donc faire preuve de l'acquisition cumulative des connaissances et des habiletés requises dans toutes les matières, telles que décrites dans *Le curriculum de l'Ontario*. Le document « Liens du TPCL avec le curriculum (2003) » récemment publié fournit quelques correspondances entre le contenu du TPCL et *Le curriculum de l'Ontario*.

Évaluation de la validité interne du test

Nous examinons les éléments suivants afin d'établir la validité interne du TPCL : a) les rapports entre chaque composante du test, b) les rapports entre les deux composantes du test, et c) la cohérence entre les réponses des élèves sur les différentes parties du test.

Les analyses qui suivent ont été basées sur des données se rapportant à des élèves récemment admissibles, soit des élèves faisant le test pour la première fois et présents les deux jours du test.

a) Rapports entre chaque composante du test

Composante lecture

Afin d'évaluer la validité interne de la composante lecture, les rapports parmi et entre les différents types de textes, d'habiletés et de questions ont été examinés. Cette composante

comprend douze textes de lecture, divisés en trois types de textes, soit i) informatifs, ii) courants avec éléments graphiques et iii) narratifs. Pour chaque type de texte, on évalue les élèves par rapport à trois habiletés en lecture, soit : i) compréhension des idées et des renseignements directement énoncés dans le texte, ii) compréhension des idées et des renseignements indirectement énoncés et iii) établissement de liens entre les connaissances et les expériences personnelles et les idées et les renseignements figurant dans les textes. Le test est composé de 40 questions à réponse choisie, 35 questions à réponse construite et 25 questions à réponse construite avec explication.

En ce qui a trait au test de langue française, les corrélations entre les trois types de textes de lecture s'échelonnent entre 0,72 et 0,78, les corrélations entre les trois habiletés en lecture s'échelonnent entre 0,76 et 0,83 et les corrélations entre les trois types d'items s'échelonnent entre 0,74 et 0,83.

En ce qui a trait au test de langue anglaise, les corrélations entre les trois types de textes de lecture s'échelonnent entre 0,73 et 0,80, les corrélations entre les trois habiletés en lecture s'échelonnent entre 0,75 et 0,81, et les corrélations entre les trois types d'items s'échelonnent entre 0,72 et 0,79.

Composante écriture

En ce qui a trait à la composante écriture, l'analyse comprend l'examen des rapports entre les quatre tâches d'écriture, et le rapport entre chacune d'elles, et le rendement global. Les quatre tâches d'écriture sont composées d'un résumé, d'un texte d'opinion en quelques paragraphes, d'une nouvelle journalistique et d'un texte informatif.

En ce qui a trait au test de langue française, les corrélations entre chacune des quatre tâches et le rendement global des élèves s'échelonnent entre 0,70 et 0,73. Ces corrélations montrent que les élèves dont le rendement est élevé sur les quatre tâches réussiront probablement la composante écriture du test. Les corrélations entre les quatre tâches d'écriture sont relativement moyennes, s'échelonnant entre 0,29 et 0,41.

En ce qui a trait au test de langue anglaise, les corrélations entre chacune des quatre tâches d'écriture et le rendement global des élèves s'échelonnent entre 0,65 et 0,69. Ces corrélations montrent que les élèves dont le rendement est élevé pour les quatre tâches réussiront sans doute la composante écriture du test. Les corrélations entre les quatre tâches d'écriture sont relativement basses, s'échelonnant entre 0,21 et 0,40.

En ce qui a trait aux résultats des tests de langue française et de langue anglaise, les corrélations relativement moyennes entre les quatre tâches d'écriture révèlent que chaque tâche d'écriture exige la connaissance d'habiletés précises en écriture qui ne sont peut-être pas identiques pour les quatre tâches. Toutefois, on s'attend à ce que les élèves aient fait l'acquisition de ces connaissances pour toutes les matières avant la fin de la 9^{e} année.

b) Rapport entre la composante lecture et la composante écriture

Le rapport global entre la composante lecture et la composante écriture était relativement élevé, avec des corrélations de 0,69 pour le test de langue française et de 0,67 pour le test de langue

anglaise. En général, les élèves dont le rendement est plus élevé pour la composante lecture ont également un rendement plus élevé en écriture. Ces résultats sont semblables aux corrélations rapportées entre les notes en lecture et en écriture (0,65) du programme d'évaluation de la 10e année à Washington (*Washington Assessment of Student Learning: Grade 10, 1999 Technical Report*, 2001).

c) Uniformité des réponses des élèves par rapport aux différentes parties de chaque composante du test

Composante lecture

Afin d'examiner l'uniformité du rendement des élèves par rapport aux trois types de textes en lecture, le rendement moyen et les écarts-types (É-T) sont présentés pour les élèves qui ont réussi la composante et pour ceux qui ne l'ont ont pas réussi.

Types de textes

i) En ce qui a trait au test de langue française, la moyenne, par rapport aux différents types de textes, des élèves qui **ont réussi** la composante lecture (n = 4 112) était comme suit :

- Textes narratifs 43 (É-T = 7,07, maximum de 60 points)
- Textes informatifs 70 (É-T = 9,44, maximum de 90 points)
- Textes courants avec éléments graphiques 41 (É-T = 5,54, maximum de 50 points)

En ce qui a trait au test de langue française, la moyenne, par rapport aux différents types de textes, des élèves qui **n'ont pas réussi** la composante lecture (n = 853) était comme suit :

- Textes narratifs 26 (É-T = 7,02, maximum de 60 points)
- Textes informatifs 44 (É-T = 9,48, maximum de 90 points)
- Textes courants avec éléments graphiques 26 (É-T = 6,51, maximum de 50 points)

ii) En ce qui a trait au test de langue anglaise, le rendement moyen, par rapport aux différents types de textes, des élèves qui **ont réussi** la composante lecture (n = 103 256) était comme suit :

- Textes narratifs 45 (É-T = 6,14, maximum de 60 points)
- Textes informatifs 64 (É-T = 8,32, maximum de 90 points)
- Textes courants avec éléments graphiques 38 (É-T = 5,18, maximum de 50 points)

En ce qui a trait au test de langue anglaise, le rendement moyen, par rapport aux différents types de textes, des élèves qui **n'ont pas réussi** la composante lecture (n = 32 733) était comme suit :

- Textes narratifs 28 (É-T = 7,67, maximum de 60 points)
- Textes informatifs 42 (É-T = 9,44, maximum de 90 points)
- Textes courants avec éléments graphiques 25 (É-T = 6,46, maximum de 50 points)

Habiletés

i) En ce qui a trait au test de langue française, le rendement moyen, par rapport aux différents types d'habiletés, des élèves qui **ont réussi** la composante lecture (n = 4 112) était comme suit :

- Comprend les idées et les renseignements directement énoncés
  50 (É-T = 6,12, maximum de 60 points)
- Comprend les idées et les renseignements indirectement énoncés
  66 (É-T = 9,53, maximum de 90 points)
- Établit des liens entre les connaissances et les expériences personnelles et les idées et les renseignements figurant dans les textes
  37 (É-T = 5,98, maximum de 50 points)

En ce qui a trait au test de langue française, le rendement moyen, par rapport aux différents types d'habiletés, des élèves qui **n'ont pas réussi** la composante lecture (n = 853) était comme suit :

- Comprend les idées et les renseignements directement énoncés
  33 (É-T = 6,70, maximum de 60 points)
- Comprend les idées et les renseignements indirectement énoncés
  40 (É-T = 8,90, maximum de 90 points)
- Établit des liens entre les connaissances et les expériences personnelles et les idées et les renseignements figurant dans les textes
  23 (É-T = 6,16, maximum de 50 points)

ii) En ce qui a trait au test de langue anglaise, le rendement moyen, par rapport aux différents types d'habiletés, des élèves qui **ont réussi** la composante lecture (n = 103 256) était comme suit :

- Comprend les idées et les renseignements directement énoncés
  47 (É-T = 5,42, maximum de 60 points)
- Comprend les idées et les renseignements indirectement énoncés
  68 (É-T = 8,20, maximum de 90 points)
- Établit des liens entre les connaissances et les expériences personnelles et les idées et les renseignements figurant dans les textes
  32 (É-T = 5,67, maximum de 50 points)

En ce qui a trait au test de langue anglaise, le rendement moyen, par rapport aux différents types d'habiletés, des élèves qui **n'ont pas réussi** la composante lecture (n = 32 733) était comme suit :

- Comprend les idées et les renseignements directement énoncés
  33 (É-T = 6,95, maximum de 60 points)
- Comprend les idées et les renseignements indirectement énoncés
  43 (É-T = 10,23, maximum de 90 points)
- Établit des liens entre les connaissances et les expériences personnelles et les idées et les renseignements figurant dans les textes
  19 (É-T = 5,82, maximum de 50 points)

Types de questions

i) En ce qui a trait au test de langue française, le rendement moyen, par rapport aux différents types de questions, des élèves **qui ont réussi** la composante lecture (n = 112) était comme suit :

- questions à réponse choisie 62 (É-T = 7,92, maximum de 80 points)
- questions à réponse construite 54 (É-T = 8,04, maximum de 70 points)
- questions à réponse construite avec explication 38 (É-T = 5,82, maximum de 50 points)

En ce qui a trait au test de langue française, le rendement moyen, par rapport aux différents types de questions, des élèves **qui n'ont pas** réussi la composante lecture (n = 853) était comme suit :

- questions à réponse choisie 42 (É-T = 7,95, maximum de 80 points)
- questions à réponse construite 31 (É-T = 8,19, maximum de 70 points)
- questions à réponse construite avec explication 22 (É-T = 6,39, maximum de 50 points)

ii) En ce qui a trait au texte de langue anglaise, le rendement moyen, par rapport aux différents types de questions, des élèves qui **ont réussi** la composante lecture (n = 103 256) était comme suit :

- questions à réponse choisie 62 (É-T = 7,23, maximum de 80 points)
- questions à réponse construite 54 (É-T = 6,68, maximum de 70 points)
- questions à réponse construite avec explication 30 (É-T = 5,82, maximum de 50 points)

En ce qui a trait au texte de langue anglaise, le rendement moyen, par rapport aux différents types de questions, des élèves qui **n'ont pas réussi** la composante lecture, (n = 32 733) était comme suit :

- questions à réponse choisie 43 (É-T = 8,43, maximum de 80 points)
- questions à réponse construite 34 (É-T = 9,32, maximum de 70 points)
- questions à réponse construite avec explication 18 (É-T = 5,74, maximum de 50 points)

En ce qui a trait aux tests de langue française et de langue anglaise, les résultats des analyses présentés sur les types de textes de lecture, d'habiletés et de questions, comparant les notes des élèves qui ont réussi la composante lecture aux notes des élèves qui ne l'ont pas réussie, ont indiqué que le premier groupe a obtenu des notes uniformément plus élevées pour les différentes parties de la composante lecture que le second groupe.

Composante écriture

Afin d'examiner l'uniformité du rendement des élèves par rapport aux quatre tâches d'écriture, les résultats sont présentés pour les élèves qui ont réussi la composante et pour les élèves qui ne l'ont pas réussie. L'échelle de notes possibles s'échelonne entre 0 et 45 points. Les résultats sont présentés en termes de mode (la note la plus fréquemment obtenue). Le pourcentage d'élèves associés au mode est présenté entre parenthèses.

i) En ce qui a trait au test de langue française, les modes des notes des différentes tâches, pour les élèves qui **ont réussi** la composante écriture (n = 4 394) étaient les suivants :

- résumé 35 (54 %)
- texte d'opinion en quelques paragraphes 35 (49 %)
- nouvelle journalistique 35 (51 %)
- texte informatif 35 (56 %)

En ce qui a trait au test de langue française, les modes des notes des différentes tâches, pour les élèves qui **n'ont pas réussi** la composante écriture (n = 571) étaient les suivants :

- résumé 0 (49 %)
- texte d'opinion en quelques paragraphes 20 (37 %)
- nouvelle journalistique 20 (32 %)
- texte informatif 0 (36 %)

ii) En ce qui a trait au test de langue anglaise, les modes des notes des différentes tâches, pour les élèves qui **ont réussi** la composante écriture (n = 115 329), étaient les suivants :

- résumé 35 (45 %)
- texte d'opinion en quelques paragraphes 35 (60 %)
- nouvelle journalistique 35 (59 %)
- texte informatif 35 (59 %)

En ce qui a trait au test de langue anglaise, les modes des notes des différentes tâches, pourles élèves qui **n'ont pas réussi** la composante écriture (n = 20 660), étaient les suivants :

- résumé 0 (49 %)
- texte d'opinion en quelques paragraphes 20 (37 %)
- nouvelle journalistique 35 (43 %)
- texte informatif 0 (50 %)

Pour les élèves de langue française et de langue anglaise qui n'ont pas réussi la composante écriture, les tâches de résumé et de texte informatif semblent avoir été les plus difficiles, comparativement aux tâches d'un texte d'opinion en quelques paragraphes et d'une nouvelle journalistique.

Pour les tests de langue française et de langue anglaise, les résultats de la comparaison des modes des notes des élèves qui ont réussi la composante écriture aux modes des notes des élèves qui ne l'ont pas réussie indiquent que le premier groupe a obtenu des notes uniformément plus élevées pour les quatre tâches de la composante écriture que le second groupe.

En ce qui a trait au test de langue anglaise, les notes pour la nouvelle journalistique ont produit le même mode pour les élèves qui ont réussi la composante écriture et les élèves qui ne l'ont pas réussie. Toutefois, le pourcentage associé aux élèves qui ont réussi la composante écriture (59 %) était plus élevé que le pourcentage associé aux élèves qui ne l'ont pas réussie (43 %).

Bibliographie

Arizona Department of Education, *Instrument to Measure Standards (AIMS) Guide to Understanding AIMS Results*. Récupéré le 19 avril 2000 de Internet à l'adresse http://www.ade.state.az.us.

Massachusetts Department of Education de (1999). *Massachusetts Comprehensive Assessment System: 1998 Technical Report* (1999). Malden : MA.

Taylor, C.S. (2001). *Washington Assessment of Student Learning: Grade 10, 1999 Technical Report.* University of Washington : Olympia.